**Words**

Which word / word groups are most mentioned? See the content of one cell as a word / word group.

**Data analysis methods**

Which statistical methods are mentioned in the list in the attachment, and how often are they mentioned? Only take statistical methods, and group together the methods that are similar.

**Words**

Of this list with biological and scientific terminology, which words relate to the field of …? Also give the counts of how many time each word / word group appears in the list.   
  
Can you take together the rows with words / word groups that are almost identical (different formats, cases, plural forms or small variations) in the … file? Make sure to update the count as well when you group together words.

In the attachment a file with words and their counts. Can you make a list of how often each single word appears in each excel sheet (so the word ‘gene expression’ counts 1 time ‘gene’ and one time ‘expression’)? Do this for every sheet separately, taking into account the frequency / count of each word group

In this list with words, how often is each word mentioned? Take each word as an individual word, so not word combinations. Take into account the frequency column as well when calculating.

I have an Excel file containing a list of terms and their associated counts. I want to group together terms that are nearly identical, accounting for:

* Differences in formatting, punctuation, and case
* Singular/plural forms (e.g., "virus" vs. "viruses", "vector" vs. "vectors")
* Common suffixes in virology like “virus,” “viruses,” “viral,” “vector,” and “vectors”

Please normalize the terms by:

1. Lowercasing all characters
2. Removing punctuation
3. Removing the suffixes listed above

For each group, return:

* representative\_term: the most representative original version from the group
* total\_count: sum of counts for all grouped terms
* grouped\_words: list of all original term variations that were grouped

I have an Excel file that contains a list of words and their frequency counts. Please help me group together terms that are nearly identical, accounting for the following:

* Noun/adjective variations (e.g., *ecology* vs *ecological*)
* Singular/plural forms (e.g., *virus* vs *viruses*)
* Morphological variants (e.g., *model*, *modeled*, *modeling*)

For each group, return:

* representative\_term: the original term from the group with the highest frequency
* total\_count: the sum of frequencies of all grouped terms
* grouped\_words: the list of all word variations included in the group

Use stemming (e.g., PorterStemmer) to group related terms. Clean any non-text entries before processing. At the end, return the result as a downloadable Excel file.